LINEAR REGRESSION – main types

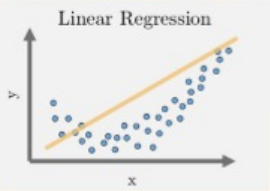
REGRESSION ANALYSIS

Fitting a function f(.) to datapoints yi=f(xi) under some error function. ie a method for estimating relationships between a dependent v. (response/outcome) and ≥1 independent variables, called 'predictors', 'covariates', 'explanatory variables', 'features'

- **simple linear regression;** one x, & one scalar y var.
- **multiple linear regression;** >1x var & one scalar y var.
- **multivariate linear regression;** multiple correlated dependent var's are predicted

LINEAR REGRESSION

Linear regression; fits the line/hyperplane by minimizing RSS, MSE; MAE error functions



$$min_\beta \sum_i \|y_i - f_\beta^{linear}(x_i)\|^2$$

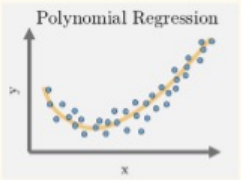
$$f_\beta^{linear}(x_i) = \beta_0 + \beta_1 x_i$$

POLYNOMIAL REGRESSION

Use n-th degree polynomial of x in its model to model non-linear x-y relationship

eg: y= w0 + w1x + w2x^2 == a statistical estimation problem is still linear (special case of multiple linear regression)

degree	polynomial equation
0	$y = a$
1	$y = ax + b$
2	$y = ax^2 + bx + c$
3	$y = ax^3 + bx^2 + cx + d$
...	...



$$min_\beta \sum_{i=0}^{m} \|y_i - f_\beta^{poly}(x_i)\|^2$$

$$f_\beta^{poly}(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_k x_i^k$$

- **higher-degree terms;** new explanatory var. resulting from the polynomial expansion of the "baseline" var. eg x^2, x^3, x^4
- **Interpretation problem;** It is often difficult to interpret the individual coeff. in a polynomial regr. fit, since the underlying monomials can be highly correlated (eg. x and x^2 may have r=0.97). It is generally more informative to consider the fitted regression function as a whole. **Point-wise or simultaneous confidence bands** can then be used to provide a sense of the uncertainty in the estimate of the regression function.
- **ALTERNATIVES;** Nonparametric regression, SVM
- **When to use ;** it is possible to approximate any continuous function on a closed interval (eg 0-100) to an arbitrary precision with a polynomial by increasing its degree.

REGRESSION WITH REGULARIZATION

TYPES & MAIN USES

Lasso — input data have only few significant predictors, that we wish to find/select to build simpler/better model

Ridge — there are many large predictors of about the same value - multicolinearity

ElasticNet — Elastic-net is useful when there are multiple features which are correlated with one another. Lasso is likely to pick one of these at random, while elastic-net is likely to pick both.

LARS — alternative to ridge and ElasticNet for multiple colinear features, iterative approach, **More effective for p>>n datasets, and alfa exploration**

LASSO REGRESSION

Can fit either a line, or polynomial minimizing the error each datapoint & the weighted L1 norm of the function parameters beta.

$$min_\beta \sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2 + \sum_{j=0}^{k} |\beta_j|$$

$$f_\beta(x_i) = f_\beta^{poly}(x_i) \text{ or } f_\beta^{linear}(x_i)$$

Used for parameter selection,

RIDGE REGRESSION

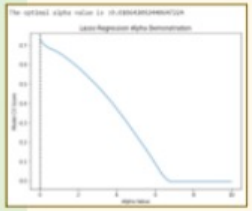
... L2 norm of the function parameters beta.

$$min_\beta \sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2 + \sum_{j=0}^{k} \beta_j^2$$

$$f_\beta(x_i) = f_\beta^{poly}(x_i) \text{ or } f_\beta^{linear}(x_i)$$

Used for estimating the coeff. of multiple-regr. models in scenarios where independent variables are highly correlated (multicollinearity)

HOT TO FIND OPTIMAL ALFA/LABDA VALUE



1.Split and Standardize the data (only model inputs and not the output)

2. Decide which regression technique Ridge, Lasso, or Elastic Net you wish to perform (or test all of them)

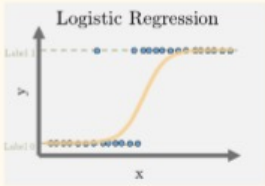
3.Use GridSearchCV to optimize the hyper-parameter alpha, and lambda ratioS in elastic net

4.Build your model with your now optimized alpha and make predictions!

GLM; Generalized linear models

LOGISTIC REGRESSION

CLASSYFICATION Can fit either a line, or polynomial with sigmoid activation f. minimizing RSS error for each datapoint. The labels y are binary class labels.



Log Regr. computes the probability of a data point being in the positive class: p(y=1|x)=sigma(fx), where f(x)=xw, is a linear regression function f(x)=xw, and sigma ∈ [0,1]. Y is a binary variable ∈{0,1}.

Weights (w) in f(x) are found by minimizing **cross-entropy (CE) loss function**, using max. likelihood estimation alg. (MLE). MLE estimates params of a likelihood function, given the observed data, to the point called maximum likelihood estimate.

Why Sigma function?; cdf of logistic distrib. has a higher kurtosis than normal .distrib. (wider tails), and cdf S shape is more flat. Thus it is used more in natural sciences

LOGREG vs PROBIT R.; probit r. uses cdf of norm. distrib. Applied more in social sc. and economy, (heteroscedastic data)

Muticlass classification; 3 approaches available

a) Softmax Regr;
b) b) OvO; one vs One,
c) c) OvR; one vs rest

Solvers; {liblinear, lbgf, SAG and SAGA},
- SAG & SAGA with large datasets,
- (L1, L2), use on scaled data

GLM; SUMMARY

Used for modelling response variables that are bounded or discrete, or to model response for var. that have large scale, and or skewed distributions.

Link Function; links the linear model to response variable

- **POISSON R; for count data.**
- **LOGISTIC & PROBIT R; for binary data**
- **MULTINOMIAL LOGISTIC & PROBIT R; for cat. data.**
- **ORDERED LOGISTIC & PROBIT R.; for ordinal data**

Summary:

	What does it fit?	Estimated function	Error Function
Linear	A line in n dimensions	$f_\beta^{linear}(x_i) = \beta_0 + \beta_1 x_i$	$\sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2$
Polynomial	A polynomial of order k	$f_\beta^{poly}(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots$	$\sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2$
Bayesian Linear	Gaussian distribution for each point	$\mathcal{N}(f_\beta(x_i), \sigma^2)$	$\sum \|y_i - \mathcal{N}(f_\beta(x_i), \sigma^2)\|^2$
Ridge	Linear/polynomial	$f_\beta^{poly}(x_i)$ or $f_\beta^{linear}(x_i)$	$\sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2 + \sum_{k=0}^{n} \beta_j^2$
LASSO	Linear/polynomial	$f_\beta^{poly}(x_i)$ or $f_\beta^{linear}(x_i)$	$\sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2 + \sum_{j=0}^{n} \|\beta_j\|$
Logistic	Linear/polynomial with sigmoid	$\sigma(f_\beta(x_i))$	$\sum_{i=0}^{m} \|y_i - f_\beta(x_i)\|^2$